

SMSSM-R 음악중재 프레임워크와 Attention-Based Multi-Task LSTM 모델을 활용한 음악 감정 분석 체계 구축

천 지 영¹ · 강 유 리² · 문 소 영³ · 윤 석 용^{4*}

¹명지대학교 AI정보과학전공 석사과정

²명지대학교 심리재활학과 박사수료

³명지대학교 음악치료학과 교수

⁴명지대학교 AI정보과학전공 교수

Development of a Music Emotion Analysis System Using the SMSSM-R Music Intervention Framework and an Attention-Based Multi-Task LSTM Model

Ji-Yeong Cheon¹ · Yu-Ri Kang² · So-Young Moon³ · Seok-Yong Yun^{4*}

¹Master's Student, Department of AI Information Science, Myongji University, Seoul 03674, Korea

²Ph.D. Candidate, Department of Psychological Rehabilitation, Myongji University, Seoul 03674, Korea

³Professor, Department of Music Therapy, Myongji University, Seoul 03674, Korea

⁴Professor, Department of AI Information Science, Myongji University, Seoul 03674, Korea

[요 약]

본 연구는 음악중재의 구조적 자극·진정 음악 모델(Structural Model for Stimulatory and Sedative Music, SMSSM)을 오디오 신호 기반으로 정량화한 수정 구조적 자극·진정 음악 모델(Revised Structural Model for Stimulatory and Sedative Music, SMSSM-R)을 제안하고, 이를 Attention-Based Multi-Task Long Short-Term Memory (LSTM) 모델과 결합하여 감정 예측과 구조적 해석 단서를 제공하는 음악 감정 분석 시스템을 구축하였다. PMemo 데이터셋을 활용한 실험 결과, SMSSM-R 변수 도입 시 Baseline 모델 대비 Valence 예측 성능의 개선 경향이 확인되었으며, 특히 통합 변수군(Case C)은 Valence와 Arousal의 주요 회귀 지표에서 가장 균형 잡힌 성능을 보였다. 이러한 결과는 SMSSM-R 기반 변수가 음악 감정 예측 모델의 입력 변수로 활용 가능하며, 예측 결과를 음악의 구조적 특성과 연결하여 해석할 수 있는 가능성을 제공함을 시사한다. 따라서 본 연구는 음악중재 이론 기반 변수를 활용한 설명 가능한 음악 감정 인식 프레임워크를 제안한다는 점에서 의의를 가진다.

[Abstract]

This study proposes the Revised Structural Model for Stimulatory and Sedative Music (SMSSM-R), which quantifies the Structural Model for Stimulatory and Sedative Music (SMSSM) framework using audio signal features. It also integrates SMSSM-R with an attention-based multi-task Long Short-Term Memory (LSTM) model to develop a music emotion analysis system that supports emotion prediction and provides structure-based interpretive cues. Experimental evaluations using the PMemo dataset showed that the SMSSM-R variables improved valence prediction compared with the baseline model. In particular, the integrated feature set (Case C) achieved the most balanced predictive performance across major regression metrics for both valence and arousal dimensions. These findings suggest that SMSSM-R-based variables can be utilized as input features for music emotion prediction and may provide interpretive cues for connecting the prediction results with the structural characteristics of music. Accordingly, this study contributes by proposing an explainable music emotion recognition framework that incorporates variables derived from music intervention theory.

색인어 : 음악 감정 인식, 딥러닝, 음악중재, SMSSM-R, Attention-Based Multi-Task LSTM

Keyword : Music Emotion Recognition, Deep Learning, Music Intervention, SMSSM-R, Attention-Based Multi-Task LSTM

<http://dx.doi.org/10.9728/dcs.2026.27.6.1647>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 15 April 2026; **Revised** 13 May 2026

Accepted 08 June 2026

***Corresponding Author; Seok-Yong Yun**

Tel:

E-mail: icanibe@mju.ac.kr

I. 서 론

현대 사회에서 음악은 단순한 예술적 감상을 넘어, 개인의 정서 상태를 조절하고 심리적 안녕을 관리하는 핵심 도구로 자리 잡고 있다. 스트리밍 서비스의 확산으로 음악 청취는 일상화되었으나, 대부분의 청취자는 음악이 자신의 감정에 미치는 영향을 객관적으로 이해하지 못한 채 직관에 의존하여 음악을 선택한다[1]. 선행 연구에 따르면, 우울 성향을 가진 청취자가 슬픈 음악을 반복적으로 청취할 경우 의도와 달리 우울감이 강화될 수 있다[2]. 이는 음악 선택이 단순한 취향의 문제가 아니라, 음악 자극과 정서 반응 간의 관계를 보다 체계적으로 이해할 필요가 있는 문제임을 시사한다.

음악 감정 인식(MER; Music Emotion Recognition) 분야는 예측 정확도 향상을 중심으로 발전해 왔으나, 예측 결과를 음악의 구조적 특성과 연결하여 해석하는 측면은 상대적으로 충분히 다루어지지 않았다. 음악중재 분야에서는 음악의 구조적 요소와 음악 경험을 임상적으로 적용하여 내담자의 정서 조절과 심리적 안녕을 증진하는 접근이 이루어져 왔다[3]. 그러나 음악중재 전문가의 임상적 통찰과 판단에 의존하는 특성상 정량화된 대규모 데이터 적용에는 한계가 있다. 두 분야가 상호 보완적인 강점을 지님에도 불구하고, 음악중재 이론에 기반한 구조적 변수를 MER 모델에 통합하고, 예측 결과를 음악적 구조와 연결하여 해석하려는 시도는 아직 충분히 이루어지지 않았다.

본 연구는 음악중재의 SMSSM(Structural Model for Stimulatory and Sedative Music) 프레임워크를 오디오 신호 처리 기반으로 정량화한 SMSSM-R(Revised)을 제안하고, 이를 MER 기술과 결합함으로써 (1) 감정 예측의 자동화, (2) 예측 결과를 음악적 구조와 연결하여 해석할 수 있는 단서 제공, (3) 대중음악 환경에서의 SMSSM 기반 변수 활용 가능성 검토를 목표로 한다.

II. 이론적 배경

2-1 음악 감정 인식(MER) 연구

1) MER의 정의와 Valence-Arousal 모델

MER은 음악 신호에서 특징을 추출하여 감정을 자동 추정하는 기술로, Russell의 Valence-Arousal(V-A) 모델에 기반하여 감정을 쾌-불쾌(Valence)와 각성-이완(Arousal)의 2차원 연속 공간으로 표현한다[4]. 리듬과 템포는 Arousal에, 음조와 화성은 Valence에 주로 영향을 미치며, 음색은 양 차원에 복합적으로 작용한다.

2) 딥러닝 기반 MER의 발전과 한계

초기 MER 연구는 SVM, K-NN, Random Forest 등 전통

적 머신러닝 기법에 의존하여 음악 감정의 시간적 전개를 충분히 반영하기 어렵다는 구조적 제약이 있었지만, TCN-LSTM-Attention 결합 모델 등 딥러닝 기반 시계열 모델 도입 이후 예측 성능이 크게 향상되었다[5]. 그러나 딥러닝 기반 MER 모델은 높은 예측 성능에도 불구하고 블랙박스적 특성을 지니며, 이는 치료적 판단이 요구되는 임상 환경에서 활용상의 제약으로 작용한다[6]. 즉 기존 MER 연구는 “어떤 감정인가(What)”에는 비교적 효과적으로 답할 수 있지만, “그 예측이 어떠한 음악적 특성과 관련되는가”를 해석하는 데는 한계가 있다.

한편 음악심리학에서는 개인의 자서전적 기억, 문화적 배경 등의 개인차가 개입되기 이전 단계에서 청취자들이 공통적으로 인지하는 구조적 정서 반응이 존재한다고 본다[7]. Gabrielsson은 이를 지각된 감정(perceived emotion)과 느껴진 감정(felt emotion)의 구분을 통해 설명하였으며[8], Eerola와 Vuoskoski는 템포, 화성, 음색, 리듬과 같은 구조적 음향 요소만으로도 Valence-Arousal 차원의 상당 부분을 설명할 수 있음을 보고하였다[7].

본 연구는 이러한 보편적 음악 지각(Universal Music Perception) 관점에 기반하여, 음악중재 전문가가 임상 현장에서 판단하는 음악 자극의 심리적 영향력을 오디오 신호 수준에서 구조화하려는 시도라고 할 수 있다.

2-2 음악중재와 SMSSM 프레임워크

1) SMSSM: 구조적 안정성과 예측 가능성 중심 모델

SMSSM(Structural Model for Stimulatory and Sedative Music)은 자극적 음악과 진정적 음악이 상이한 정서적·생리적 반응을 유도한다는 이론에 근거하여 음악의 구조적 안정성과 예측 가능성을 중심으로 음악을 분석하는 프레임워크이다[9]. Grocke의 SMMA와 Wigram의 PSSM을 통합하여 개발되었으며, 스타일/형식, 박자, 리듬, 조성, 선율, 화성, 음색, 음량, 긴장도 등 13개 상위 범주 43개 하위 요소로 음악을 평가한다[10],[11]. SMSSM 점수가 높을수록 구조적 안정성이 높아 진정 효과가 강해지며, 낮을수록 자극 효과를 유발하는 것으로 해석한다.

2) SMSSM 보완의 필요성과 본 연구의 접근

SMSSM은 충분한 내용 타당도를 지니지만, 전문가의 수기 평정에 의존하여 실시간 분석 및 대규모 데이터 적용이 어렵고 평정자 간 일치도 문제를 내포한다. 또한 이론적 기반인 SMMA가 클래식 음악 분석을 전제로 설계된 만큼, 리듬 중심의 반복적 화성 구조를 가진 현대 대중음악에 동일하게 적용 가능한지에 대한 실증 검증이 필요하다[10],[11].

기존 음악과 정서를 다룬 연구에서는 음악의 정서적 효과를 구조적으로 설명하는 이론이 제시되어 왔으나, 자동화와 확장성에는 한계가 있다. 반면 MER 연구는 높은 예측 성능을

보이지만 예측 결과에 대한 해석 가능성이 부족하다. 두 접근이 상호 보완적 강점을 지님에도 불구하고, 이를 통합한 설명 가능한 프레임워크는 아직 충분히 제시되지 않았다.

한편, 본 연구에서 활용한 일부 변수는 장조·단조, 화성 안정성 등 서양 음악 이론의 영향을 포함하고 있어 비서구권 음악이나 특수 장르에 대한 적용 가능성을 추가적으로 검토할 필요가 있다. 그러나 템포, 리듬의 규칙성, 음량 변화와 같은 주요 변수들은 특정 문화권에 국한된 이론 개념이라기보다 인간의 생리적·지각적 반응과 연관된 음향심리학적 특성(acoustic universals)에 기반한다. 선행 연구 또한 템포와 리듬 구조가 문화권을 초월하여 인간의 집단행동 및 정서 반응과 관련된 보편성을 가질 수 있음을 보고하였다[12]. 이에 SMSSM-R은 템포, 리듬, 음량 변화, 음색 등 오디오 신호 기반의 다양한 음향 특징을 활용하여 청취자가 지각하는 음악의 구조적·정서적 특성을 반영하도록 설계되었다.

이에 본 연구는 SMSSM을 오디오 신호 처리 기반으로 정량화한 SMSSM-R과 MER 기술을 결합하여, 감정 예측의 자동화와 음악적 구조에 기반한 해석 가능성을 동시에 제공하는 음악 감정 분석 체계를 제안한다.

III. 연구 방법

3-1 PMEmo 데이터셋

본 연구는 연속적 감정 주석과 생체 신호를 제공하는 PMEmo 데이터셋의 대중음악 794곡을 활용하였다[13]. PMEmo는 청취자가 실제로 경험한 유발된 감정(induced emotion)을 0.5초 단위로 기록했다는 점에서 음악중재 및 정서 조절 연구에 적합하다[14]. 데이터는 총 401명의 응답을 기반으로 구축되었으며, Cronbach's $\alpha=0.998$ 로 높은 주석 신뢰도를 보인다. 감정 주석과 오디오 특징의 정합이 가능한 곡을 선별하고, 초기 청취 적응 효과를 줄이기 위해 재생 시작 후 15초 이전 구간을 제외하여 최종 767곡을 EDA 및 변수 생성 대상으로 사용하였다. 다만 PMEmo의 연속 자기보고 방식은 청취자의 인지적 처리 지연, 주관적 보정, 개인차의 영향을 완전히 배제하기 어렵다. 본 연구에서는 EDA 단계에서 시차 기반 상관 분석을 수행하고, 이를 바탕으로 감정 반응 지연을 부분적으로 보정하였다.

3-2 SMSSM-R 정량화 체계

1) 개요

본 연구는 음악치료 박사 수료 이상의 전문가 3인과 논의하여 기존 SMSSM 구성 요소 중 오디오 신호로 정량화 가능한 요소를 선별하였다. 그 결과 SMSSM-R은 Librosa 라이브러리를 통해 자동 추출 가능한 총 54개 변수로 구성되었다:

Valence 설명변수 11개(V1~V11), Arousal 설명변수 11개(A1~A11), 진정-자극 구조 변수 32개. 모든 변수는 0~1 범위로 정규화하였다.

변수 구성의 학술적 타당성을 검토하기 위해 음악치료 박사 수료 이상의 전문가 10인을 대상으로 2차에 걸친 델파이(Delphi) 조사를 실시하였다. 주요 검토 항목은 ① SMSSM 이론적 속성의 반영도, ② Librosa 기반 음향 특징 연산 방식의 타당성, ③ 자극 방향 변수의 역산 정렬 방식과 음악중재적 해석 기준의 부합성이었다.

각 변수의 수식, 가중치, 임계값 및 정규화 상수는 SMSSM 원 이론과 음악심리학 선행 연구를 바탕으로 설정된 초기 설계값(initial heuristic parameters)에 해당하며, 향후 데이터 기반 최적화 및 ablation study를 통한 추가 검증이 필요하다.

2) Valence 설명 변수(V1~V11)

Valence 설명변수는 음악의 구조적 안정성, 조성적 일관성, 선율의 매끄러움 등 긍정 정서 유발 요인을 정량화하도록 설계하였다.

① V1. 형식적 균형성(Structural Balance): RMSE, Spectral Centroid, MFCC, Chroma를 결합한 특징 벡터 기반 자기유사도 행렬(SSM)을 구성하여 음악 구조의 반복성과 통일감을 정량화하였다[15].

$$V1 = \overline{S_{off}} \times C_{90} \quad (1)$$

② V2. 박자의 규칙성(Metrical Regularity): Onset Strength의 국소 최댓값을 비트로 간주하고, 비트 간 시간 간격의 표준편차를 이용하여 박자의 시간적 일관성을 산출하였다.

$$V2 = \frac{1}{\sigma(\Delta t_i) + \epsilon} \quad (2)$$

③ V3. 리듬의 유연성(Rhythmic Fluidity): Onset Strength의 프레임 간 변화량과 비트 간 시간 간격의 변동성을 결합하여 리듬 흐름의 안정성을 측정하였다.

$$V3 = \frac{1}{\sigma(\Delta O) + \sigma(\Delta t_{beat}) + \epsilon} \quad (3)$$

④ V4. 템포의 적절성(Tempo Fitness): 실제 템포와 onset 밀도 기반 암시적 기대 템포 간 부합도 및 구간 내 템포 안정성을 가중 결합하였다.

$$V4 = 0.7 \times F_{fit} + 0.3 \times F_{stable} \quad (4)$$

⑤ V5. 조성적 색채(Tonal Color): Krumhansl-Schmuckler 키 프로파일과 Chroma 벡터 간 상관도를 기반으로 조성 강도($T_{strength}$)와 장조 성향을 산출하였다.

$$V5 = T_{strength} \cdot \log(1 + B_{mode}) \quad (5)$$

⑥ V6. 선율의 심미성(Melodic Aesthetics): MIDI 음고 기반 음정 간격을 이용하여 선율 진행의 매끄러움(S), 반복 진행 비율(R_{rep}), 순차진행비율(R_{step})을 반영하였다.

$$V6 = 0.5 \times S + 0.3 \times (1 - R_{rep}) + 0.2 \times R_{step} \quad (6)$$

⑦ V7. 화성적 협화도(Harmonic Consonance): Spectral Centroid 안정성, Chroma 엔트로피($\overline{H}_{chroma}$), Spectral Flatness(SF) 및 Tonal Centroid 변화량을 결합하여 화성 진행의 안정성을 산출하였다.

$$V7 = 0.6 \times C_{harm} + 0.4 \times (1 - \tanh(\overline{H_{CDF}})) \quad (7)$$

⑧ V8. 음색의 온화함(Timbral Warmth): Spectral Centroid의 윈도우 평균을 kHz 단위로 변환한 뒤 역수를 취하여 저음역 중심의 따뜻한 음색 특성을 측정하였다.

$$V8 = \frac{1}{\mu_t(SC)/1000 + \epsilon} \quad (8)$$

⑨ V9. 다이내믹 적정성 (Dynamic Adequacy): RMS 에너지로부터 Integrated Loudness(L_{cur})와 Loudness Range(LRA_{cur})를 산출하고, 각각 ITU-R BS.1770 기준값 (-14 LUFS, 8 LU)를 비교하여 산출하였다[16],[17].

$$V9 = 0.6 S_L + 0.4 S_{LRA} \quad (9)$$

⑩ V10. 긴장의 해소 (Musical Resolution): Chroma Flux를 이용하여 안정 구간의 점유율을 산출하고, 지속적 긴장 해소 상태를 반영하기 위해 제곱 가중을 적용하였다[15].

$$V10 = R_{stable}^2 \quad (10)$$

⑪ V11. 음향적 명료도(Acoustic Clarity): Spectral Flatness(SF_t)의 보완 지표 ($1 - Flatness$)를 이용해 순간 명료도(C_t) 시계열을 구성하고, 윈도우 내 평균 수준과 시간적 일관성 ($1 - \sigma_C$)을 가중 결합하였다.

$$V11 = 0.7 \times \mu_t(C_t) + 0.3 \times (1 - \sigma_C) \quad (11)$$

3) Arousal 설명 변수(A1~A11)

Arousal은 주로 템포, 리듬, 음량, 고음역 에너지에 의해 영향을 받는다. Arousal 설명변수는 에너지 수준, 리듬 복잡도, 화성적 긴장감 등 청취자의 생리적·심리적 각성 반응을 유발하는 음악 구조적 요인을 정량화하도록 설계하였다.

① A1. 형식적 복잡성(Formal Complexity): Spectral Flux 기반 novelty curve에서 임계값을 초과하는 국소 최대값의 밀도를 산출하여 구조 전환 빈도를 측정하였다[15].

$$A1 = \frac{N_{peaks}}{T_{window}} \quad (12)$$

② A2. 박자의 가변성(Metric Variability): Gaussian 평활화된 Onset Strength의 변동성과 비트 간격의 변동계수를 결합하였다.

$$A2 = 0.5 \cdot \sigma(\hat{O}) + 0.5 \cdot CV_{beat} \quad (13)$$

③ A3. 리듬 밀도&당김음(Rhythmic Density & Syncopation): 윈도우 내 onset 밀도와 정규화된 PLP(Predominant Local Pulse)의 분산을 결합하여 리듬 이벤트의 빈도와 펄스 불안정성을 측정하였다.

$$A3 = \frac{N_{onset}}{T_{window}} \cdot (1 + \sigma \cdot Var(\hat{P}_{norm})) \quad (14)$$

④ A4. 템포의 가변성(Tempo Variability): Beat tracking으로 산출한 BPM의 전역 분산과 평활화된 국소 BPM 변화율을 가중 결합하여 속도 변화의 역동성을 측정하였다.

$$A4 = 0.6 \cdot \min\left(\sigma \frac{(BPM)}{50}, 1\right) + 0.4 \cdot \min(\mu(|\nabla|), 1) \quad (15)$$

⑤ A5. 조성적 긴장감(Tonal Tension): Krumhansl-Schmuckler 키 프로파일과의 최대 상관도(ρ_{max})로 으뜸음 지배력의 약화를 측정하고, 장·단조 프로파일 상관도 차이 ($|\rho_M^* - \rho_m^*|$)를 이용해 조성 중심의 불명확성과 선법 모호성을 측정하였다[18].

$$A5 = 0.6 \cdot (1 - \rho_{max}) + 0.4 \cdot (1 - |\rho_M^* - \rho_m^*|) \quad (16)$$

⑥ A6. 선율의 도약성(Melodic Leapiness): 기본 주파수 (F0)를 MIDI 음고로 변환한 뒤, 음역 폭과 상위 25% 음정 도약의 빈도·크기를 결합하였다.

$$A6 = Range_{\pi th} \times \frac{N_{\leq ap}}{N_{total}} \times \frac{\mu_{\leq ap}}{12} \quad (17)$$

⑦ A7. 화성적 불협화도(Harmonic Dissonance): 7개 주파수 대역별 Spectral Contrast의 시간적 변동성(표준편차)을 산출하고 그 평균을 취하여 배음 충돌에 따른 화성적 거칠기를 산출하였다.

$$A7 = \frac{1}{B} \sum_{b=1}^B \sigma(C_b) \quad (18)$$

⑧ A8. 음색 밝기&날카로움(Timbral Brightness): Spectral Centroid의 윈도우 평균으로 밝기, Spectral Rolloff 95%의 윈도우 평균으로 날카로움을 산출하였다.

$$A8_{brightness} = \mu_t(SC) \quad (19)$$

$$A8_{sharpness} = \mu_t(SR_{95}) \quad (20)$$

⑨ A9. 다이내믹 강도(Dynamic Intensity): RMS 에너지 기반 다이내믹 레인지와 변동계수를 결합하고 tanh 함수를 적용해 음량 변화의 역동성을 측정하였다.

$$A9 = \tanh\left(\frac{\max(S) - \min(S)}{\mu_S + \epsilon} \times \frac{\sigma_S}{\mu_S + \epsilon}\right) \quad (21)$$

⑩ A10. 긴장-이완 대비(Tension-Relaxation Contrast): Chroma Flux, RMSE, Onset Strength를 정규화한 뒤 균등 가중 합산하여 통합 긴장도 시계열을 구성하고, 정점-기저 대비를 평균 긴장도로 정규화하였다[15].

$$A10 = \frac{\max(T_t) - \min(T_t)}{\mu(T_t) + \epsilon} \quad (22)$$

⑪ A11. 현장감&선명도(Presence & Clarity): 정규화된 Spectral Bandwidth, Rolloff 95%, 크레스트 팩터 프록시를 가중 결합하였다.

$$A11 = 0.4 \cdot \frac{\mu_t(BW)}{f_{Nyquist}} + 0.3 \cdot \frac{\mu_t(Rolloff)}{f_{Nyquist}} + 0.3 \cdot (1 - \mu_t(Flatness)) \quad (23)$$

4) 진정-자극 설명 변수(32개)

진정-자극 설명 변수는 SMSSM 이론에서 제시하는 구조적 안정성, 반복성, 예측 가능성 개념을 오디오 신호 기반으로 정량화한 변수 집합이다. 형식, 박자, 리듬, 템포, 조성, 선율, 화성, 음색, 다이내믹, 긴장-이완, 음질의 11개 범주에서 총 32개 변수를 설계하였다. 복잡성, 변화 빈도, 변동성 등 자극 방향을 나타내는 변수에는 부호 반전을 적용하여, 값이 높을수록 구조적 안정성 또는 진정적 특성이 크게 나타나도록 하였다. 또한 모든 변수에 Min-Max 정규화(0~1)를 적용하였다.

본 연구의 진정-자극 변수는 긍정·부정 정서나 선호도를 직접 측정하기 위한 지표가 아니라, 음악 구조의 안정성, 반복성, 예측 가능성에 기반한 자극 수준을 정량화하기 위한 변수이다. 따라서 반복성이 높고 변화가 적은 음악은 구조적으로 진정적 특성을 가진 것으로 해석할 수 있으나, 이것이 반드시 긍정적 정서나 높은 선호도를 의미하는 것은 아니다.

구체적으로 형식 범주에서는 Chroma 기반 자기유사성 행렬(SSM)을 활용하여 구조적 반복성과 변화 수준을 측정하였다. 리듬 범주에서는 Zero-Crossing Rate 변동성과 Onset Strength 기반 SSM을 이용해 리듬의 규칙성과 반복성을 반영하였다. 선율 범주에서는 F0를 MIDI 음고로 변환한 후 반음 단위 변화량을 산출하였으며, 화성 범주에서는 Chroma Entropy와 Chroma Flux를 활용하여 화성적 분산과 변화성을 측정하였다. 다이내믹 범주에서는 RMSE 기반 음량 변화 특성을, 긴장-이완 범주에서는 Chroma Flux를 이용한 긴장 축적 및 해소 관련 변화를 반영하였다. 음질 범주에서는 Spectral Flatness, MIDI 기반 음고 분산, Chroma 집중도 등을 활용하여 음향적 안정성과 명료성 관련 특성을 측정하였다. 생성된 변수와 설명은 표 1과 같다.

$$SMSSM-R_{sedative} = \begin{cases} X, & \text{if } X \uparrow \Rightarrow \text{calm} \\ 1-X, & \text{if } X \uparrow \Rightarrow \text{stimulative} \end{cases} \quad (24)$$

3-3 탐색적 데이터 분석(EDA) 및 전처리

1) 데이터 정합성 확보 및 전처리

767곡을 대상으로 EDA를 수행하였으며, music_id를 기준으로 Metadata, Annotation, Audio 데이터를 통합하였다. 원시 오디오에서 Librosa를 활용하여 Spectral Centroid, Zero-Crossing Rate, MFCC, Chroma, RMS Energy 등 저수준 음향 특징을 추출하였다[19]. 음향 특징은 약 23ms 단위 프레임으로 계산되므로, 0.5초 단위 감정 주석과의 시간 해상도 정합을 위해 Window-Based Averaging을 적용하였다. 또한 데이터셋의 주석 시작 시점을 고려해 재생 후 15초 이전 구간은 분석에서 제외하였다. 이후 SMSSM-R 기반 변수 생성을 통해 V/A 설명변수 22개, 진정-자극 구조 변수 32개를 도출하였다. 자극 방향 변수는 부호 반전 처리를 통해 진정 방향으로 통일하였으며, BPM 표준편차와 F0 기반 선율 변수에는 상한 클리핑을 적용하였다. 최종적으로 모든 입력 변수에 Min-Max 정규화(0~1)를 적용하였다.

2) 데이터 구조 및 품질 점검

전처리된 Master 데이터셋은 총 36,434개 프레임과 54개의 SMSSM-R 기반 음향 특징 변수로 구성되었다. 종속 변수는 Valence와 Arousal이다. Baseline 데이터셋은 Librosa 기반 저수준 음향 특징 67개로 별도 구성하였으며, Case C는 다중공선성 제거 후 최종 36개 변수로 구성하였다. 입력 변수의 분포 확인 결과, 리듬·템포·구조 변화 관련 변수에서 강한 양의 왜도와 다수의 이상치가 관찰되었다. 이에 따라 상·하위 퍼센타일 기반 클리핑 후 Min-Max 정규화를 수행하였다. Valence-Arousal 공간에서 두 감정 차원 모두 중·고 영역에 집중되는 경향이 확인되었으며, K-means 군집분석에서 Elbow method, Silhouette 분석, 군집 균형을 종합하여 K=5를 최종 군집 수로 선정하였다. 이러한 결과는 PMEmo

표 1. SMSSM-R 기반 진정-자극 정량 변수 정의

Table 1. Definition of sedative-stimulative quantitative features based on SMSSM-revised

Category	Variable Name	Definition
Form	Formal_Complexity	Inverse measure of structural complexity reflecting simplicity and comprehensibility of musical form
	Frequency_of_Structural_Change	Inverse measure of structural change frequency indicating structural stability
Repetition	Repetition_of_Musical_Elements	Degree of repetition in melody, rhythm, and harmony reflecting familiarity and predictability
Meter	Metric_Variability	Inverse measure of metrical variability reflecting temporal regularity
Silence	Frequency_of_Silence_Pauses	Inverse measure of silence occurrence reflecting continuity of sound flow
	Predictability_of_Silence_Pauses	Predictability of pause timing reflecting stability of musical phrasing
Rhythm	Rhythmic_Regularity_Periodicity	Degree of rhythmic regularity and periodic repetition
	Repetition_of_Rhythmic_Patterns	Degree of repeated rhythmic patterns reflecting familiarity
	Rhythmic_Variability	Inverse measure of rhythmic variation reflecting rhythmic stability
	Predictability_of_Rhythm_Change	Predictability of rhythmic transitions and changes
Tempo	Tempo_Variability	Inverse measure of tempo variation reflecting tempo stability
	Predictability_of_Tempo_Change	Predictability of tempo transitions within expected ranges
Tonality	Frequency_of_Key_Change	Inverse measure of tonal change frequency reflecting tonal stability
	Predictability_of_Key_Change	Predictability of tonal modulation and harmonic movement
Melody	Frequency_of_Melodic_Change	Inverse measure of melodic variation reflecting melodic stability
	Predictability_of_Melody	Predictability of melodic progression and contour
Harmony	Consonance_Dissonance_Level	Degree of harmonic consonance relative to dissonance
	Harmonic_Density_Complexity	Inverse measure of harmonic density and complexity
	Frequency_of_Harmonic_Change	Inverse measure of harmonic progression changes
	Predictability_of_Harmonic_Progression	Predictability of chord progression and harmonic flow
Timbre	Timbre_Instrumentation_Variability	Inverse measure of timbral and instrumental variability
Dynamics	Dynamic_Range	Inverse measure of loudness variation reflecting dynamic stability
	Impression_Change	Inverse measure of expressive variation reflecting consistency of musical impression
	Frequency_of_Dynamic_Change	Inverse measure of dynamic change frequency
	Predictability_of_Dynamic_Change	Predictability of dynamic transitions
Tension/Relaxation	Tension_Building	Inverse measure of tension accumulation reflecting relaxation tendency
	Relaxation	Degree of tension release and relaxation
	Clarity_of_Resolution	Clarity and completeness of tension resolution
	Delay_Lack_of_Resolution	Inverse measure of delayed tension resolution
Quality	Recording_Quality	Clarity and fidelity of audio recording quality
	Performance_Quality	Stability and consistency of musical performance
	Emotional_Clarity	Clarity of intended emotional expression

데이터셋이 극단적 저각성·저정서 구간보다 중·고 감정 영역에 더 많은 표본을 포함하고 있음을 보여준다. 따라서 본 연구 모델의 예측 결과는 중·고 감정 영역에서 상대적으로 안정적으로 해석될 수 있으나, 극단적 저각성·저정서 영역에 대한 일반화에는 한계가 존재한다.

스케일링 방법 선택을 위해 Standard, Min-Max, Robust 스케일링을 비교한 결과, Min-Max 스케일링이 군집 간 표본수의 균형 면에서 가장 우수하여 최종 적용하였다.

3) 다중공선성 및 시차 특성 분석

SMSSM-R 기반 54개 변수 간 다중공선성을 점검하기 위해 분산팽창계수(VIF)를 산출하였다. 완전 선형 종속으로

VIF=∞인 변수 3개를 먼저 제거한 뒤, 잔여 변수의 VIF가 15 미만이 될 때까지 최댓값 변수를 순차 제거하였다. 그 결과 초기 54개 중 18개가 제거되어 최종 36개 변수가 Case C 입력 변수로 선정되었다. 일반적으로 VIF > 10 기준이 널리 활용되지만, 본 연구에서 해당 기준을 적용할 경우 최종 변수가 14개로 축소되어 리듬, 화성, 긴장-이완 등 일부 음악 구조 범주의 대표성이 과도하게 저하되는 문제가 확인되었다. 따라서 본 연구에서는 변수 대표성과 다중공선성 완화 간의 균형을 고려하여 VIF < 15를 절충 기준으로 적용하였다[20]. 다만 이 기준은 방법론적 선택에 해당하므로, 향후 더 큰 표본과 다양한 데이터셋을 통한 추가 검증이 필요하다.

시차 기반 Cross-correlation 분석 결과, Arousal 반응성

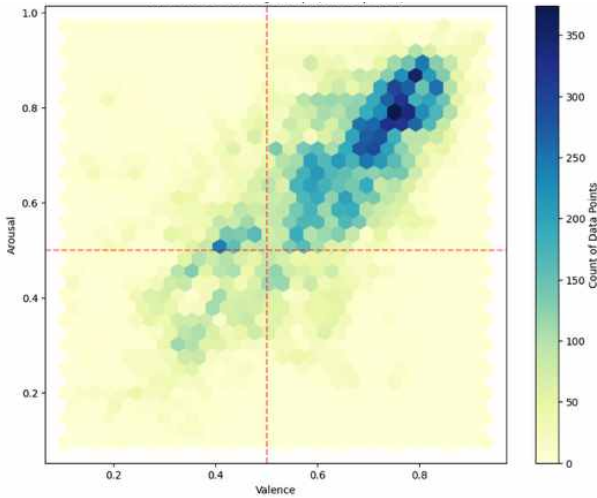


그림 1. Valence-arousal 공간에서의 감정 분포

Fig. 1. Emotion distribution in the valence-arousal space

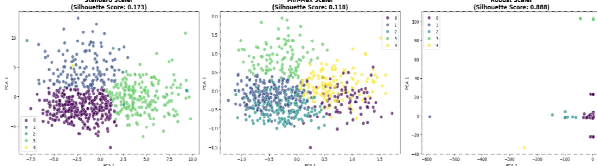


그림 2. 스케일링 방법에 따른 PCA 기반 군집 분포 비교

Fig. 2. PCA-based cluster distribution by scaling method

상위 변수들의 최적 시차는 5~6프레임, 즉 약 2.5~3초에 집중되어 각성 반응의 지연 특성이 확인되었다. Valence 반응성 상위 변수들의 최적 시차는 0~8프레임 범위에 분산되었으나, 두 차원의 시차 중앙값이 모두 5프레임으로 나타났다. 본 연구에서는 방법론적 일관성을 위해 Valence와 Arousal 모두에 5프레임(약 2.5초)의 지연 보정을 적용하였다. 이에 따라 지연 보정 및 슬라이딩 윈도우 기반 시퀀스 구성 과정에서 유효 시퀀스 생성이 어려운 일부 프레임 및 곡이 제외되었으며, 최종 분석 데이터셋은 751곡, 32,627프레임으로 구성되었다.

표 2. 데이터 전처리, 변수 선택 및 분석 데이터 개요

Table 2. Overview of data preprocessing, feature selection, and analysis dataset

Item	Processing Method
Non-normality and Outliers	Upper and lower percentile-based clipping followed by Min-Max normalization (0~1)
Emotional Response Delay	Delay correction of 5 frames (approximately 2.5 seconds) applied to both Valence and Arousal
Multicollinearity	Sequential removal based on VIF < 15 criterion (18 variables removed; 36 variables retained)
Data Splitting	Split at the music_id level to prevent frames from the same track being distributed across different subsets
Final Analysis Dataset	751 tracks and 32,627 frames

4) EDA 종합 및 전처리 전략

본 연구는 SMSSM-R 이론 기반 변수의 효과를 단계적으로 검증하기 위해 네 가지 실험 케이스를 구성하였다. 모든 케이스에는 동일한 Attention-Based Multi-Task LSTM 모델을 적용하여, 성능 차이가 모델 구조가 아니라 입력 변수 구성의 차이에서 기인하도록 통제하였다. Baseline은 Librosa 라이브러리를 통해 추출한 저수준 음향 특징 67개로 구성하였다.

3-4 실험 설계

표 3. 실험 케이스 구성

Table 3. Configuration of experimental cases

Case	Feature Composition	Number of Features	Theoretical Basis
Baseline	Librosa acoustic features	67	Pure low-level acoustic signal representation
Case A	SMSSM-Revised (Valence-Arousal)	22	Direct explanatory variables for Valence-Arousal dimensions
Case B	SMSSM-Revised (Sedative-Stimulative)	32	Music therapy-based structural framework
Case C	Integrated features (after VIF removal)	36	Multidimensional musical characteristics with reduced multicollinearity

1) **데이터 분할**: music_id 기준으로 train 64%, validation 16%, test 20% 비율로 분할

2) **반복 실험**: Repeated Random Holdout, random seed 5개(42~46), 최종 성능은 mean ± std로 보고

3) **하이퍼파라미터 결정**: learning rate {0.01, 0.001, 0.0001}, hidden dimension {32, 64, 128}, dropout {0.1, 0.2, 0.3}에 대한 Grid Search (27개 조합)

표 4. 케이스별 최적 하이퍼파라미터

Table 4. Optimal Hyperparameters by Case

Case	Learning rate	Hidden dim	Dropout
Baseline	0.001	128	0.3
Case A	0.01	32	0.3
Case B	0.01	128	0.3
Case C	0.01	128	0.3

4) **학습 환경**: Apple MacBook Pro(Apple M4, 2024), Apple Silicon MPS 백엔드, Python 3.11.15, PyTorch 2.11, 배치 크기 32, 최대 에폭 20, Adam 옵티마이저 (weight decay=1e-5)

5) 검증 가설:

① SMSSM-R 기반 변수(Case A, B, C)는 Baseline 대비 높은 감정 예측 성능을 보일 것이다.

② V/A 설명변수(Case A)와 진정-자극 변수(Case B)는

Valence 예측에서 서로 다른 성능 특성을 보일 것이다.

③ 통합 변수(Case C)는 Valence와 Arousal 전반에서 가장 균형 잡힌 예측 성능을 보일 것이다.

3-5 모델 아키텍처

음악 감정은 시간에 따라 연속적으로 변화하며, EDA 결과에서 Valence와 Arousal 모두 시차 반응 특성이 확인되었다. 이에 본 연구는 Bidirectional LSTM, Self-Attention, Multi-Task 출력 헤드를 결합한 Attention-Based Multi-Task LSTM을 설계하였다. 본 연구에서는 슬라이딩 윈도우로 구성된 입력 시퀀스 내에서 양방향 문맥을 학습하도록 설계하였으며, 곡 전체의 미래 정보를 활용하는 구조는 아니다. 또한 Self-Attention 층을 추가하여 입력 시퀀스 내에서 감정 예측에 중요한 시간 구간의 정보를 효과적으로 반영할 수 있도록 하였다. Valence와 Arousal은 동일한 감정 공간을 구성하지만 서로 다른 반응 특성을 가진다. Valence는 조성, 화성, 선율 구조와 같은 인지적·해석적 요인과 관련되는 반면, Arousal은 템포, 음량, 리듬 밀도 등 물리적 에너지 특성에 상대적으로 민감하다. 이에 두 감정 차원이 공통 시계열 표현을 공유하면서도 최종 출력 단계에서는 각각 독립적인 완전연결층을 통해 예측되도록 Multi-Task 구조를 적용하였다. 입력은 슬라이딩 윈도우 방식으로 구성하였다. 윈도우 크기는 20프레임(약 10초)이며, 각 입력 시퀀스는 현재 시점 기준 이전 20개 프레임의 음향 특징으로 이루어진다. 타깃 값은 마지막 프레임의 Valence와 Arousal이다. 모델은 2개의 Bidirectional LSTM 레이어와 Self-Attention 층으로 구성되며, 이후 각 감정 차원을 예측하는 독립적인 완전연결층을 연결하였다. Bidirectional 구조에 따라 최종 은닉 표현의 차원은 hidden dimension의 2배가 된다.

3-6 평가 지표

Valence와 Arousal은 서로 다른 반응 특성을 지니므로, 본 연구는 두 감정 차원을 독립적인 회귀 지표로 평가하였다. 단일 통합 지표만 사용할 경우 한 차원의 성능 향상이 다른 차원의 오류를 상쇄할 수 있기 때문이다. 주 평가 지표로는 Pearson 상관계수(r), 평균절대오차(MAE), 결정계수(R^2)

를 사용하였다. 또, Valence-Arousal 2차원 공간을 각 축의 중간값(0.5) 기준 4사분면으로 구분한 Quadrant Accuracy를 추가 산출하였다.

IV. 연구 결과

4-1 케이스별 성능 비교 및 해석 가능성 분석

1) 예측 성능 평가 결과

예측 성능은 music_id 기준 5회 반복 무작위 분할 실험의 평균 및 표준편차(mean $\pm$ std)로 평가하였다. SMSSM-R 기반 변수군을 사용한 Case A, B, C는 전반적으로 Baseline 대비 향상된 성능을 보였다. 특히 Case C는 Valence와 Arousal의 Pearson 상관계수 및 Arousal의 R^2 에서 가장 높은 성능을 나타내어 두 감정 차원의 연속값 예측에 가장 효과적인 것으로 확인되었다. 반면 Quadrant Accuracy는 Case A가 가장 높게 나타나 감정 사분면 구분에 상대적 강점을 보였다.

2) Permutation Importance 분석

각 입력 변수를 무작위로 치환한 후 Pearson r 감소량을

표 6. 실험 케이스별 permutation importance 상위 5개 변수 (Δr mean 기준)

#	Baseline	Case A	Case B	Case C
1	MFCC_1 (0.153)	a9_dynamic_intensity (0.050)	Timbre_Inst_Variability (0.188)	a9_dynamic_intensity (0.071)
2	Spectral_BW (0.094)	a2_meter_variability (0.025)	Harmonic_Density (0.100)	Timbre_Inst_Variability (0.049)
3	Spectral_Roll off (0.041)	v7_harmonic_consonance (0.023)	Impression_Change (0.058)	v8_timbre_warmth (0.015)
4	Spectral_Centroid (0.024)	a8_timbral_brightness (0.010)	Freq_Dynamic_Change (0.036)	v7_harmonic_consonance (0.013)
5	MFCC_3 (0.015)	v8_timbre_warmth (0.007)	Pred_Dynamic_Change (0.035)	a4_tempo_variability (0.007)

표 5. 실험 케이스별 예측 성능 평가 결과 (Mean $\pm$ std, $n=5$)

Table 5. Comparative performance of experimental cases (Mean $\pm$ std, $n=5$)

Case	Pearson r (V)	Pearson r (A)	MAE (V)	MAE (A)	RMSE (V)	RMSE (A)	R^2 (V)	R^2 (A)	Quadrant Acc
Baseline	0.5543 $\pm$ 0.0638	0.6799 $\pm$ 0.0352	0.1069 $\pm$ 0.0065	0.1035 $\pm$ 0.0047	0.1355 $\pm$ 0.0104	0.1322 $\pm$ 0.0066	0.2849 $\pm$ 0.1019	0.4469 $\pm$ 0.0444	0.6846 $\pm$ 0.0357
Case A	0.6483 $\pm$ 0.0368	0.7016 $\pm$ 0.0114	0.1014 $\pm$ 0.0075	0.1008 $\pm$ 0.0021	0.1290 $\pm$ 0.0080	0.1295 $\pm$ 0.0036	0.3915 $\pm$ 0.0544	0.4738 $\pm$ 0.0221	0.7124 $\pm$ 0.0150
Case B	0.6192 $\pm$ 0.0580	0.6852 $\pm$ 0.0413	0.1057 $\pm$ 0.0094	0.1053 $\pm$ 0.0060	0.1335 $\pm$ 0.0118	0.1323 $\pm$ 0.0067	0.3447 $\pm$ 0.1045	0.4492 $\pm$ 0.0584	0.6821 $\pm$ 0.0240
Case C	0.6575 $\pm$ 0.0355	0.7457 $\pm$ 0.0234	0.1004 $\pm$ 0.0086	0.0926 $\pm$ 0.0046	0.1286 $\pm$ 0.0125	0.1203 $\pm$ 0.0067	0.3973 $\pm$ 0.0582	0.5463 $\pm$ 0.0327	0.7066 $\pm$ 0.0215

계산하여 변수 중요도를 산출하였다. Baseline에서는 MFCC와 Spectral Bandwidth 등 저수준 음향 특징의 중요도가 높게 나타났다. 반면 SMSSM-R 기반 케이스에서는 개별 변수의 중요도가 전반적으로 낮게 나타났으며, 이는 복합 구조 변수의 정보가 여러 변수에 분산되어 있기 때문으로 해석된다. 또한 Case C에서는 Case A와 Case B의 주요 변수들이 함께 상위 변수에 포함되어 통합 변수군의 활용 가능성을 확인할 수 있었다. 다만 Permutation Importance 값은 변수 구성, 변수 간 정보 중복성, 스케일링 방식의 영향을 받을 수 있으므로, 본 연구에서는 절댓값 비교보다 각 케이스 내부에서 어떤 유형의 변수가 예측에 상대적으로 크게 기여했는지를 중심으로 해석하였다.

3) Attention 가중치 분석

Attention 가중치 분석 결과, Baseline은 입력 윈도우 후반부에 높은 가중치가 집중된 반면, SMSSM-R 기반 케이스는 가중치가 윈도우 전반에 분산되어 나타났다. 이는 SMSSM-R 기반 변수가 특정 시점의 저수준 신호 변화에만 의존하기보다, 입력 윈도우 전반의 구조적 정보를 활용했을 가능성을 시사한다. 다만 Attention 가중치는 감정 유발 원인에 대한 인과적 설명이 아니라, 모델 내부에서 학습된 상대적 가중치를 나타내는 보조적 해석 지표로 보아야 한다.

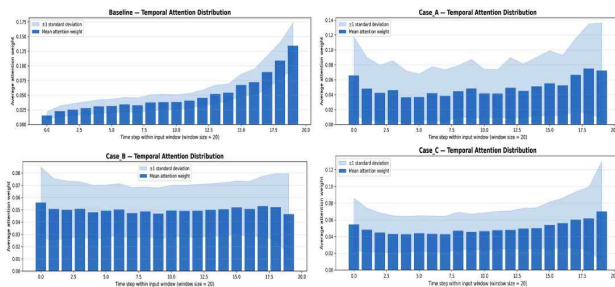


그림 3. 케이스별 temporal attention 가중치 분포

Fig. 3. Temporal attention weight distribution by case

4) 통계적 유의성 검증

Baseline 대비 성능 차이를 검토하기 위해 대응표본 t-검정을 수행하였다. 그 결과, Valence Pearson r은 모든 SMSSM-R 기반 케이스에서 Baseline 대비 통계적으로 유

의한 차이를 보였으며, Arousal Pearson r은 Case C에서만 유의한 차이가 확인되었다. R^2 의 경우 Valence에서는 Case A와 Case C가, Arousal에서는 Case C가 Baseline 대비 통계적으로 유의한 차이를 보였다. 다만 반복 실험 수가 5회로 제한되어 있으므로, 통계적 유의성 결과는 탐색적 검증으로 해석할 필요가 있다. 향후 반복 실험 수를 확대하고 외부 데이터셋을 활용한 추가 검증이 요구된다(* $p < 0.05$, ** $p < 0.01$, ns: not significant).

4-2 가설 검증

1) SMSSM-R 기반 변수의 Baseline 대비 효과: 가설 ①

SMSSM-R 기반 모든 케이스는 Valence 예측에서 Baseline 대비 유의한 성능 향상을 보였다($p < 0.05$). Arousal 예측에서는 Case C에서만 유의한 향상이 확인되었다. 따라서 가설 ①은 부분적으로 지지되었다.

2) V/A 설명변수와 진정-자극 구조 변수의 역할 차이: 가설 ②

Valence Pearson r 기준으로 Case A는 Case B보다 평균 값이 높게 나타났으나, 표준편차 범위가 상당 부분 중첩되어 Case A의 명확한 우위가 입증되었다고 보기는 어렵다. 그러나 Case A는 Valence 예측에서 상대적으로 높은 평균 성능을 보였고, Case B 역시 Baseline보다 높은 Valence Pearson r을 보여 진정-자극 구조 변수도 Valence 예측에 일정 부분 기여할 가능성이 확인되었다. 따라서 가설 ②는 부분적으로 지지되었다.

3) 통합 변수군의 균형적 성능: 가설 ③

Case C는 Valence와 Arousal의 Pearson r 및 R^2 에서 모두 가장 높은 평균 성능을 보였다. Baseline 대비 t-검정에서도 Valence Pearson r, Arousal Pearson r, Valence R^2 , Arousal R^2 에서 통계적으로 유의한 향상을 보였다. 이는 통합 변수군이 두 감정 차원을 비교적 균형 있게 예측했음을 시사한다. 따라서 가설 ③은 부분적으로 지지되었다.

4-3 종합 해석

본 실험의 핵심 결과는 다음 세 가지로 요약된다.

첫째, SMSSM-R 기반 변수는 Baseline 대비 Valence 예측에서 일관된 개선 경향을 보였으며, Arousal 예측에서는 통합 변수군인 Case C에서 가장 뚜렷한 개선이 확인되었다. 이는 음악중재 이론 기반 변수 설계가 음악 감정 예측에 활용될 수 있음을 시사한다.

둘째, V/A 설명변수와 진정-자극 구조 변수는 Valence 예측에서 일부 중첩된 설명 가능성을 보였다. Case B도 Baseline보다 높은 Valence Pearson r을 보여, 협화적 화성 구조·규칙적 리듬·템포 패턴 등 SMSSM의 구조적 안정성 변수가 생리적 이완 반응뿐 아니라 인지적·정서적 평가 차원과

표 7. Baseline 대비 케이스별 t-test 결과

Table 7. Paired t-test results: Baseline vs each case

Metrics	Case A	Case B	Case C
Valence Pearson r	$\Delta=+0.094$, $p=0.027^*$	$\Delta=+0.065$, $p=0.044^*$	$\Delta=+0.103$, $p=0.011^*$
Arousal Pearson r	$\Delta=+0.022$, $p=0.307$ ns	$\Delta=+0.005$, $p=0.775$ ns	$\Delta=+0.066$, $p=0.017^*$
Valence R^2	$\Delta=+0.107$, $p=0.038^*$	$\Delta=+0.060$, $p=0.252$ ns	$\Delta=+0.112$, $p=0.016^*$
Arousal R^2	$\Delta=+0.027$, $p=0.178$ ns	$\Delta=+0.002$, $p=0.912$ ns	$\Delta=+0.099$, $p=0.008^{**}$

도 관련될 수 있음을 시사한다.

셋째, VIF 기반으로 변수 중복성을 완화한 Case C는 대부분의 회귀 지표에서 가장 높은 성능을 보였다. 이는 단순히 변수 수를 늘리는 것보다, 이론적으로 구분된 변수군을 통합 하되 다중공선성을 완화하는 방식이 두 감정 차원의 균형적 예측에 유리할 수 있음을 시사한다.

단, 본 연구의 해석 가능성 분석은 감정 유발 원인을 직접 규명하는 인과적 설명이 아니라, 예측 결과를 음악중재 이론 기반 변수와 연결하여 해석할 수 있는 보조적 단서 제공 수준임을 명시한다. 케이스 간 성능 차이의 절대적 크기가 크지 않고 반복 실험 수가 제한적이므로, 본 결과는 후속 연구를 통해 검증해야 할 가능성으로 해석하는 것이 적절하다.

V. 결 론

본 연구는 음악중재의 SMSSM 프레임워크를 오디오 신호 처리 기반으로 정량화한 SMSSM-R을 제안하고, 이를 Attention-Based Multi-Task LSTM과 결합한 음악 감정 분석 체계를 구축하였다. PMEmo 데이터셋을 대상으로 네 가지 실험 케이스를 비교한 결과, SMSSM-R 기반 변수는 저 수준 음향 특징만을 사용한 Baseline 대비 Valence 예측에서 일관된 개선 경향을 보였다. 또한 VIF 기반 다중공선성 제거를 거친 통합 변수군(Case C)은 Valence와 Arousal의 주요 회귀 지표에서 가장 균형 잡힌 성능을 보였다. 이는 음악 중재 이론 기반 변수가 MER 모델의 입력 변수로 활용될 수 있으며, 예측 결과를 음악 구조와 연결하여 해석할 수 있는 보조적 단서를 제공할 수 있음을 시사한다. 본 연구에는 다음과 같은 한계가 있으며, 이에 따른 향후 연구 방향은 다음과 같다.

(1) SMSSM-R 변수의 가중치, 임계값, 정규화 상수는 이론적 근거와 전문가 검토를 기반으로 설정되었으나, 데이터 기반 최적화나 전문가 평정값과의 직접 비교 검증은 충분히 수행되지 않았다. 따라서 향후 연구에서는 전문가 평정 데이터와의 비교 검증, 민감도 분석, ablation study를 통해 SMSSM-R 변수 수식과 가중치, 임계값의 타당성을 추가로 검증할 필요가 있다.

(2) 본 연구는 Librosa 기반 full-mix 오디오 특징을 사용하였기 때문에 보컬, 리듬, 화성 악기 등 개별 음원 수준의 정밀한 분석에는 한계가 있다. 따라서 향후 연구에서는 Demucs, Spleeter와 같은 소스 분리 기법을 적용하여 선율, 화성, 리듬 특징을 개별 음원 수준에서 보다 정교하게 추출할 필요가 있다.

(3) 연구 대상이 PMEmo 데이터셋의 대중음악에 한정되어 있어 클래식, 재즈, 국악 및 비서구권 음악에 대한 일반화 가능성은 추가 검증이 필요하다. 본 연구의 핵심 변수들이 음향심리학적 보편성(acoustic universals)에 근거하여 다양한

장르의 정서 에너지를 포착할 수 있도록 설계되었으나, 향후 연구에서는 국악을 포함한 비서구권 전통 음악, 클래식, 재즈, 실험적 장르의 데이터를 확충하여 SMSSM-R 프레임워크의 범용적 타당성을 다각적으로 검토할 필요가 있다.

(4) 데이터셋의 감정 분포가 중·고 Valence 및 중·고 Arousal 영역에 집중되어 있어 극단적 저각성·저정서 영역에 대한 예측 신뢰도는 제한적일 수 있다. 또한 진정-자극 변수는 구조적 안정성과 반복성이 높을수록 진정적 특성이 증가한다는 가정에 기반하지만, 실제 음악 반응은 지루함이나 과도한 단조로움과 같은 비선형적 반응을 포함할 수 있다[21]. 따라서 향후 연구에서는 극단적 감정 영역의 표본을 보완하고, 역U자형 반응 곡선 등 비선형 감정 모델을 변수 설계에 반영할 필요가 있다.

(5) 본 연구는 음악 자극의 구조적·음향적 특성에 기반한 감정 예측에 초점을 두었기 때문에, 개인의 기억, 문화적 배경, 음악 선호도와 같은 비음향적 요소는 모델에 포함하지 못하였다. 음악 감정은 본질적으로 개인 경험과 사회문화적 맥락의 영향을 받는 복합적 현상이므로, 본 연구는 개인차가 개입되기 이전 단계의 '보편적 음악 지각' 수준을 중심으로 해석하는 데 한정된다. 따라서 본 모델은 음악 자극 자체의 구조적 정서 특성을 해석하기 위한 베이스라인 프레임워크로 이해하는 것이 타당하다. 향후 연구에서는 자서전적 기억, 선호도, 문화적 배경, 생체신호 등을 결합한 멀티모달(multimodal) 접근을 통해 보다 개인화된 음악 감정 예측 모델로 확장할 필요가 있다.

이러한 한계에도 불구하고, 본 연구는 다음과 같은 의의를 가진다. 첫째, 전문가 수기 평정에 의존하던 SMSSM의 주요 개념을 오디오 신호 기반 변수로 정량화하여 대규모 음악 데이터 분석에 적용 가능한 형태로 확장하였다. 둘째, Valence-Arousal 설명변수와 진정-자극 구조 변수를 비교·통합함으로써 음악중재 이론 기반 변수가 MER 모델에서 활용될 가능성을 실험적으로 검토하였다. 셋째, Permutation Importance와 Attention 가중치 분석을 통해 모델 예측 결과를 음악 구조 기반 변수와 연결하여 해석할 수 있는 가능성을 제시하였다. 결론적으로 본 연구는 음악중재 이론과 딥러닝 기반 MER을 연결하는 설명 가능한 음악 감정 인식 프레임워크의 기초를 제시했다는 점에서 의의를 가진다.

참고문헌

- [1] A. van Goethem and J. Sloboda, "The Functions of Music for Affect Regulation," *Musicae Scientiae*, Vol. 15, No. 2, pp. 208-228, 2011. <https://doi.org/10.1177/1029864911401174>
- [2] S. Garrido and E. Schubert, "Music and People with Tendencies to Depression," *Music Perception*, Vol. 32, No. 4, pp. 313-321, 2015. <https://doi.org/10.1525/mp.2015.32.4>

- [3] K. E. Bruscia, *Defining Music Therapy*, 3rd ed. University Park, IL: Barcelona Publishers, 2014.
- [4] X. Cui, Y. Wu, J. Wu, Z. You, J. Xiahou, and M. Ouyang, "A Review: Music-Emotion Recognition and Analysis Based on EEG Signals," *Frontiers in Neuroinformatics*, Vol. 16, 997282, 2022. <https://doi.org/10.3389/fninf.2022.997282>
- [5] D. Han, Y. Kong, J. Han, and G. Wang, "A Survey of Music Emotion Recognition," *Frontiers of Computer Science*, Vol. 16, 166335, January 2022. <https://doi.org/10.1007/s11704-021-0569-4>
- [6] F. Doshi-Velez and B. Kim, "Towards a Rigorous Science of Interpretable Machine Learning," arXiv:1702.08608, 2017. <https://doi.org/10.48550/arXiv.1702.08608>
- [7] T. Eerola and J. K. Vuoskoski, "A Comparison of the Discrete and Dimensional Models of Emotion in Music," *Psychology of Music*, Vol. 39, No. 1, pp. 18-49, 2011. <https://doi.org/10.1177/0305735610362821>
- [8] A. Gabrielsson, "Emotion Perceived and Emotion Felt: Same or Different?," *Musicae Scientiae*, Vol. 5, No. 1_suppl, pp. 123-147, 2001. <https://doi.org/10.1177/10298649020050S105>
- [9] J. Kim, S. Moon, H. Ko, M. Choi, and J. Park, "Music Analysis Applied to an Integrative Arts Therapy - Focusing on Sacred Circle Dance Music," *Journal of Arts Psychotherapy*, Vol. 15, No. 3, pp. 313-333, 2019. <https://doi.org/10.32451/KJOAPS.2019.15.3.313>
- [10] D. Grocke and T. Wigram, *Receptive Methods in Music Therapy: Techniques and Clinical Applications for Music Therapy Clinicians, Educators and Students*, London, UK: Jessica Kingsley Publishers, 2006.
- [11] T. Wigram, *Improvisation: Methods and Techniques for Music Therapy Clinicians, Educators, and Students*, London, UK: Jessica Kingsley Publishers, 2004.
- [12] P. E. Savage, S. Brown, E. Sakai, and T. E. Currie, "Statistical Universals Reveal the Structures and Functions of Human Music," *Proceedings of the National Academy of Sciences*, Vol. 112, No. 29, pp. 8987-8992, 2015. <https://doi.org/10.1073/pnas.1414495112>
- [13] K. Zhang, H. Zhang, S. Li, C. Yang, and L. Sun, "The PMEmo Dataset For Music Emotion Recognition," in *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval (ICMR)*, Yokohama, Japan, pp. 135-142, June 2018. <https://doi.org/10.1145/3206025.3206037>
- [14] J. Lee, Emotional Regulation Functions of Music Listening, Ph.D. Dissertation, Seoul National University, Seoul, February 2015. <https://doi.org/10.23170/snu.0000000026243.11032.0000434>
- [15] J. Foote, "Automatic Audio Segmentation Using a Measure of Audio Novelty," in *Proceedings of the 2000 IEEE International Conference on Multimedia and Expo (ICME)*, New York: NY, pp. 452-455, July 2000. <https://doi.org/10.1109/ICME.2000.869637>
- [16] International Telecommunication Union, Algorithms to Measure Audio Programme Loudness and True-Peak Audio Level, ITU, Geneva, Switzerland, Recommendation ITU-R BS.1770-4, 2015. <https://www.itu.int/rec/R-REC-B.1770/en>
- [17] European Broadcasting Union, Loudness Normalisation and Permitted Maximum Level of Audio Signals, EBU, Geneva, Switzerland, Recommendation EBU R 128, 2020.
- [18] D. Temperley, "What's Key for Key? The Krumhansl-Schmuckler Key-Finding Algorithm Reconsidered," *Music Perception*, Vol. 17, No. 1, pp. 65-100, 1999. <https://doi.org/10.2307/40285812>
- [19] B. McFee, C. Raffel, D. Liang, D. P. Ellis, M. McVicar, E. Battenberg, and O. Nieto, "Librosa: Audio and Music Signal Analysis in Python," in *Proceedings of the 14th Python in Science Conference (SciPy)*, Austin, TX, pp. 18-24, 2015. <https://doi.org/10.25080/Majora-7b98e3ed-003>
- [20] R. M. O'Brien, "A Caution Regarding Rules of Thumb for Variance Inflation Factors," *Quality & Quantity*, Vol. 41, No. 5, pp. 673-690, 2007. <https://doi.org/10.1007/s11135-006-9018-6>
- [21] D. E. Berlyne, *Aesthetics and Psychobiology*, New York, NY: Appleton-Century-Crofts, 1971.



천지영(Ji-Yeong Cheon)

2025년 : 명지대학교 정치외교학과
(학사)

2025년 ~ 현 재: 명지대학교 기록정보과학전문대학원
AI정보과학전공 석사과정

※관심분야 : AI Agents, RAG, 데이터분석, LLM/sLLM, 딥러닝



강유리(Yu-Ri Kang)

2008년 2월 : 명지대학교
사회교육대학원
음악치료학과
(음악치료학 석사)

2025년 9월 : 명지대학교 대학원
심리재활학과 박사수료

※관심분야 : 음악치료, 심리재활, 수용음악치료



문소영(So-Young Moon)

2000년 : 호주 The University of
Melbourne (M. Mus.
음악치료 전공)

2008년 : 호주 The University of
Melbourne (Ph. D. 음악치료
전공)

2015년 ~ 현 재: 명지대학교 통합치료대학원 음악치료학과
교수

2022년 ~ 현 재: World Congress of Music Therapy
Scientific Committee Member

2023년 ~ 현 재: 한국통합치료학회 감사

2024년 ~ 현 재: 한국자연치유학회 편집부위원장

2025년 ~ 현 재: 대한통합의학교육협회 음악치료분과 이사

※관심분야 : 음악치료, 신경재활, 수용음악치료



윤석용(Seok-Yong Yun)

2000년 : 숭실대학교 정보과학대학원
정보산업학과 (이학석사)

2018년 : 숭실대학교 일반대학원
IT정책경영학과 (공학박사)

1990년 ~ 1994년: (주)포스코 전산시스템부 사원

1994년 ~ 1999년: (주)현대정보기술 IT건설팀 팀 책임

2000년 ~ 2015년: (주)포스코경영연구원 빅데이터TFT 부장

2015년 ~ 2020년: (주)베가스 빅데이터건설팀 부사장

2019년 ~ 현 재: 명지대학교 기록정보과학전문대학원

AI정보과학전공 주임교수

※관심분야 : 데이터분석, 기계학습, 인공지능, 자연어처리,
LLM/sLLM, 데이터베이스, 데이터 거버넌스,
CDS